贝叶斯网络在适应性教学系统中的应用研究

Applied Research of Bayes Net in Adaptive CAI

马文龙 (衢州学院信息与电子工程系 浙江衢州 324000)
(浙江师范大学计算机科学研究所 浙江金华 321004)
瞿有甜 (浙江师范大学计算机科学研究所 浙江金华 321004)
张金伟 (衢州学院信息与电子工程系 浙江衢州 324000)
(浙江师范大学计算机科学研究所 浙江金华 321004)

摘 要: 适应性教学系统中学生模型不确定性因素的处理是非常重要的问题,本文运用了贝叶斯网络来处理这些不确定性的因素。文章重点分析了学生模型中贝叶斯网络的构造过程,并应用一种查找聚类结点的算法对贝叶斯网络进行优化。

关键词: 贝叶斯网络 学生模型 适应性教学系统

1 引言

当前,适应性教学系统已成为教育技术领域备受关注的研究内容。适应性教学是指为了提高学习的适应性,通过学习环境的自身调整来适应学习者的个性特征和认知水平,满足学习者个别需要的学习过程。适应性教学系统主要特点是系统能够根据学习者的学习特征提供与其相适应的学习方式和学习内容,真正实现个性化的教学。一个典型的适应性教学系统由专家知识、学生模型、教师模型和人机接口等模块组成。学生模型是为系统在教学中能动态分析学生特征而设计的,它用来记录学生的个人情况、学习目标、学习进展情况和学习掌握水平等信息,力求反映学生对知识掌握情况的准确信息,以便教师模型中的教学策略部分能够及时准确的做出决断。因而,学生模型是实施适应性教学的基础,如何建立一个准确、高效的学生模型是适应性教学系统设计的核心任务。

设计一个学生模型难点在于处理学生学习过程中存在的众多不确定性因素[1],如关于学生领域知识的掌握程度的不确定性;学生的浏览动作对知识掌握程度贡献不确定性;学生的浏览动作和目标联系的不确定性等等。这些不确定性可以引入贝叶斯网络进行推理,本文将重点讲述笔者所设计和实现的适应性教学系统中如何采用贝叶斯网络,处理学生模型中众多的不确定性因素,建立一个覆盖型的学生模型。

2 贝叶斯网络

贝叶斯网络是一种概率推理机制[2-3],它是在概率论的基础上进行不确定推理。贝叶斯网络为在某一特定应用领域中描述随机变量之间的概率依赖性提供了一个图形化的表达方式,以及利用这些依赖进行复杂的概率推理的算法。

定义 1:贝叶斯网络

设 $V=\{X_1,X_2,\cdots\cdots,X_n\}$ 是值域 U 上的 n 个随机变量,则值域 U 上的贝叶斯网络 $BN(B_S,B_P)$,其中:

(1) $B_S=(V,E)$ 是一个定义在 V 上的有向无环图 Γ(DAG),V 是该有向无环图 Γ 的节点集,E 是 Γ 的边集。如果存在一条节点 X_i 到节点 X_j 的有向边,则称 X_i 是 X_j 的父节点,X_j 是 X_i 的子节点。记 X_i 的所有父节点为 πX_i。

(2) $B_p=\{P(X_i|\pi X_i)[0,1]|X_i\in V\}$。对于 V 中的每个节点,定义了一组条件概率分布函数 $P(X_i|\pi X_i)[0,1]$。

即:给定一个有向无环图 Γ 和一个离散变量集合 $X=\{X_1,X_2,\cdots\cdots,X_n\}$ 上的联合概率分布 P,如果 Γ 可以代表 P,即在 X 中的变量和 Γ 的节点之间存在一一对应的关系,使得 P 可以进行如下的递归乘积分解,如

公式(1)所示:

$$P(X_1)\prod_i P(X_i \mid \pi X_i) \quad 式(1)$$

这里 πX_i 是图 Γ 中 X_i 的直接祖先(父节点)。我们将图 Γ 和概率分布 P 的联合称为贝叶斯网络。

一个贝叶斯网络由网络结构表示其定性部分,由条件概率分布函数表示其定量部分。除了对域进行定义,这两部分必须加以指明以构成一个贝叶斯网络,之后在一个基于知识的系统中被用作推导引擎。

3 学生模型中贝叶斯网络构造

从信息处理的角度分析,建立学生模型的实质是分析和处理学生的相关信息,包括知识的表示和学生认知诊断两个问题。知识的表示我们采用覆盖型学生模型,学生认知诊断采用贝叶斯网络进行推理。

目前有许多建立学生模型的方法和理论[4],常见的学生模型有覆盖型、偏差型和认知型。覆盖型是描述学生知道什么、不知道什么的学生模型,它把学生的知识描述成领域知识模块中专业知识的一个子集。即设一个领域知识以树形来表示,整个领域知识树为 DKT,设学生已掌握的知识单元构成的知识树为 GKT,未掌握的知识树为 GNKT。则有等式:{GNKT} = {DKT}—{GKT};如果{GNKT}为"空"时,则表示该学生已达到了各知识单元的全部要求。如果{GNKT}不为"空",则该学生未掌握的知识单元可从 GNKT 中获得。覆盖模型的优点是可以作为有大量知识点的课程测试中的学生模型,能很清楚的表示先验知识。基于以上优点,本文采用了覆盖型学生模型。

认知的诊断是判断学生对该知识点的掌握情况。学生模型知识表示的网络结构复杂性及学生学习过程中存在的众多不确定性因素,决定了认知诊断的复杂性。为此我们将贝叶斯网络引入学生建模中,利用其对于先验信息和后验信息的结合能力,对学生所学知识项之间的依赖关系进行推理。由公式(1)可以得到公式(2)所示的贝叶斯规则。

$$P(C \mid S) = \frac{P(C) * P(S \mid C)}{P(S)} \quad 式(2)$$

设定 C 事件表示学生知道某个概念,S 事件表示学生实施了某个动作(如求解步骤),P(C)表示学生知道某个概念的先验概率,P(S|C)是学生在知道某个概念的前提下能够完成某个求解步骤的概率,由此公式,学生的动作可以演绎出对某个概念的掌握程度。

将贝叶斯网络应用于适应性教学系统有一个前提是必须拥有足够多的条件数据。本文采用一种人工和自动相结合的方式构造贝叶斯网络。先由课程的领域专家——授课老师根据其课堂教学的经验,给出该课程知识点的先验概率,系统再进行领域知识建模、生成贝叶斯网络结构、确定条件概率分布函数。在系统投入使用一定时期并收集到足够多的学生数据后最后进行贝叶斯网络推理和优化。

3.1 领域知识建模和网络结构的确定

我们建立的学生模型要模拟知识在学生头脑中的存储。领域之间的内在逻辑关系把某门课程中的知识连接成一个语义网络。学生在学习过程中也是按照这种逻辑顺序进行学习的。因此,我们需将知识间的逻辑关系首先描述清楚,即领域知识的建模。具体实现过程:

(1) 划分基本知识项。适应性教学系统的每一门课程都是基于知识项(KI, Knowledge Item)[5]这个概念,知识项不同于传统的知识点,它是课程中信息的一种抽象,反映了课程内部的语义关系。为了定义一门课程中的相关知识项,首先我们按照粒度层次表示方法由领域专家把一门课程基本知识按照由粗到细顺序逐级分解,最终成为基本的知识项,如《C 语言程序设计》中数据类型、常量和变量等可成为基本的知识项。基本知识项下面由若干测试项来验证该知识点是否掌握。

(2) 确定知识项之间的依赖关系。领域专家将知识项间的依赖关系主要区分为聚合和先决两种。聚合关系表示了一种知识间的包含关系,它用来将较概括的概念分解为较基本的概念,表现为一种树型的分支结构,如程序的基本结构与顺序结构、选择结构、循环结构知识点之间就形成了聚合关系。先决关系则表示了知识间逻辑上的顺序关系,即在掌握某知识点前,必须掌握另一知识点作为基础。这种关系表现为在不同的树型分支间的路径。$KI_1 < KI_2$ 表示 KI1 是理解 KI_2 的前序知识,必须在学习 KI_2 之前学习。例如《C 语言程序设计》中为了理解指针变量首先要了解变量的概念。

(3) 划分层次。为了便于贝叶斯网络进行推理,还需按照知识项之间的依赖关系进行划分层次,具体分为 3 层。第 1 层(level1)包括最简单的概念和知识,

掌握它们不需要理解前序的知识项；第 2 层((Level2)是较高一级的知识，想要掌握它们就需要理解第一层的某些知识项；第 3 层((Level3)比第 2 层知识项更难懂，它们的掌握需要有第 1、2 层知识项作前序知识项。同一层的知识项之间是相互独立的，没有先决关系；不同层次之间的主要知识项之间有先决关系。然后将第 2 层的知识项进一步划分成两类 Level 2 和 Level 2´。Level 2 的知识项是 Level 3 的前序知识项；Level 2' 级所包含的知识项都不是 Level 3 的前序知识项。

这样就建立了一种相对规范化的课程表示方法，将一门课程知识项按依赖关系连接在一起就形成了一张知识项依赖图，如图 1 所示，图 1 中'⃠'表示测试项结点，'○'表示知识项结点，'↖'表示聚合关系，'↙'表示先决关系。这张依赖图即为贝叶斯的网络结构。

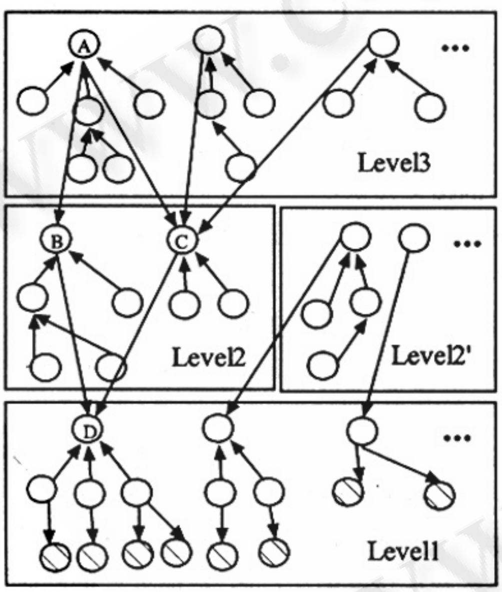


图 1 知识项依赖图

3.2 条件概率分布函数的确定

领域专家给每个知识项定义一张条件概率分布表，用于量化知识项间学习依赖关系的强弱程度。在适应性教学系统中，我们考虑图 1 中的每个知识项结点为二值节点，其中表示两个状态：掌握(True)、未掌握(False)；用概率 P(K1 = True)度量学生对知识项 K1 的理解程度，概率 P(K1 = False)度量学生对知识项 K1 的陌生程度。显然有，P(K1 = True) + P(K1 = False) = 1。

贝叶斯网络进行推理之前首先要确定推理路径方向，方向表示了知识项间的因果或先后顺序，我们可以根据知识项间的依赖关系确定这种推理方向。对于知识间的聚合关系，采用向上聚集路径方向。领域专家认为“当学生掌握了次一级的所有概念，则其必然掌握总的知识概念”。先决关系是从复杂的概念推向作为基础的简单概念，它的路径方向应该按照由高级概念到低级概念的方向。我们根据各个结点的入度和出度情况，将图中所有的结点分为四类，然后根据每个节点所属类别规定条件概率推理方式。具体分类如下：

(1) 只有聚合关系指向的节点。如果有 n 个分支，其中 r 个掌握了，则此条件概率设为 r/n。

(2) 同时有聚合关系和先决关系指向的节点。当表示先决关系的路径的父节点为 False 时，条件概率与只有聚合关系指向的节点算法相同，即不考虑先决关系。当有表示先决关系的路径的父节点为 True 时，无论有几个，为条件概率加 50%，然后根据其他 n 个聚合关系的父节点中 r 个已掌握，则最终条件概率取 50% + r/2n。

(3) 只有先决关系指向的节点。父节点任一一个为 True，条件概率都为 90%；当所有先决关系的父节点都为 False 时，条件概率为 30%。

(4) 测试项节点。测试项节点条件概率应考虑失误率 m 和猜中率 g，失误率指用户答错自己知道的问题的概率，猜中率指用户答对自己不知道问题的概率。如果考虑上述两上因素，则条件概率如表 1 所示：其中 K 表示测试项的父结点，即基本的知识点；Q 表示测试项结点。

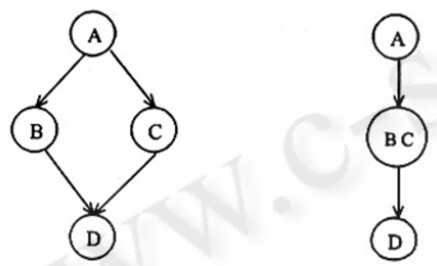
表 1 测试项结点条件概率表

P(Q|K)	P(K = True)	P(K = False)
P(Q = True)	1 - m	g
P(Q = False)	m	1 - g

3.3 贝叶斯网络推理和优化

贝叶斯网络在已知某些结点概率值的情况下，利用贝叶斯网络计算可以获得其他结点的条件概率，这种推理可以形象的称为条件概率的“传播”。当贝叶斯网络中不存在无向环的结构时，可以找到线性算法，

如公式(1)和公式(2)所示。在适应性教学系统中，知识项之间划分了层次，每个层次内不存在无向环，但level1 结点、Level2 结点和 Level3 结点之间由于知识项相关度较高，存在许多无向环，如图 1 中节点 A－D 就形成一个无向环。由于存在无向环的贝叶斯推理是一个 NP－Hard 问题[6]，必须找到一种算法消除无向环，优化贝叶斯网络的结构。通常可以采用聚类(Clustering)算法。如图 2 所示，将图中 B 和 C 合并为一个结点，从而消除图 2(a)中的无向环，但如何找到正确的聚类结点是该算法的关键。我们利用系统收集的数据，经过计算和分析，找到了一种算法来找到相应的聚类结点。



(a) 包含无向环的 BN　(b) 经过聚簇后的 BN

图 2　聚类处理

算法：查找聚类结点

输入：贝叶斯网络图 G＝{V,E}

输出：聚类的结点序列 C1,C2,…Cn

处理：

(1) i＝1

(2) 寻找在图 G 中入度＝0 的结点 v，将 v 加入到 Ci。

(3) $\forall v \in V$，且 $v \notin C_i$，对于 $\forall u \in C_i$，如果边 $\{u,v\} \in E$，且将 v 加入到 Ci 中不形成无向环，则将 v 加入 Ci。

(4) 重复步骤(3)，直到无法找到满足条件(3)中的条件 v，即得结点序列 Ci。

(5) 将图 G 删除 Ci 所对应的边和结点得到新图 G'。

(6) 如果 G'不为空，令 G＝G'，i＝i＋1，跳到步骤(2)。

采用上述算法处理，优化后的贝叶斯网络结构如图 3 所示。

4　结束语

适应性教学系统建立学生模型的过程中存在许多不确定性的因素，而贝叶斯网络正好为其提供了一种有效的推理机制，其难点是如何针对适应性教学系统的特点构造一个有效、准确的贝叶斯网络。本文着重

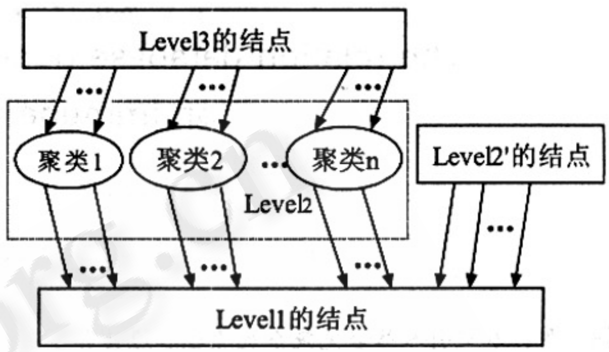


图 3　优化后的贝叶斯网络

研究了该贝叶斯网络的构造过程，并应用聚类算法对贝叶斯网络进行优化。但是建立准确的贝叶斯网络是有一定难度的，有许多细节需要继续研究。

参考文献

1 Jameson A. Numerical uncertainty management in user and student modeling: an overview of systems and issues[J]. User Modeling and User－Adapted Interaction, 1995.5.

2 Kevin Patrick Murphy. "A Brief Introduction to Graphical Models and Bayesian Networks". Department of Computer Science at U. C. Bakeley, 10, 2000.

3 Judea Pearl. Bayesian Cognitive Systems Laboratory Computer Science. Department University of California Los Angeles, 1994.

4 杨国才、徐要学，一个智能教学系统的设计模型，计算机应用，1998,10:17－19.

5 谢深泉，知识点及其网络的特性分析[J]，软件学报，1998(10):50－53.

6 G Cooper. The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks[J]. Artificial Intelligence, 1990:42(2－3):393－405.